



Uso de escalado automático en AWS (Linux)



INTRODUCCIÓN

la interfaz de línea de comandos de AWS (AWS CLI) permite crear una instancia de Amazon Elastic Compute Cloud (EC2) para alojar un servidor web y crear una imagen de Amazon Machine (AMI) a partir de esa instancia. AWS AMI permite iniciar un sistema que se escala automáticamente bajo una carga variable mediante el uso de Amazon EC2 Auto Scaling.

OBJETIVOS

- Cree una instancia EC2 mediante un comando AWS CLI.
- Cree una nueva AMI mediante la AWS CLI.
- Cree una plantilla de lanzamiento de Amazon EC2.
- Cree una configuración de lanzamiento de Amazon EC2 Auto Scaling.
- Configure políticas de escalamiento y cree un grupo de escalamiento automático para ampliar o reducir la cantidad de servidores en función de una carga variable.



TAREA 1

En esta tarea, se inicia una nueva instancia de EC2 y, luego, se crea una nueva AMI basada en esa instancia en ejecución. Se utiliza la CLI de AWS en la instancia de EC2 del host de comandos para realizar todas estas operaciones.

- En la Consola de administración de AWS, en la barra de búsqueda, ingrese y elija EC2 abrir la Consola de administración de Amazon EC2.
- En el panel de navegación, seleccione Instancias.
- De la lista de instancias, seleccione la Instancia de host de comando.

The screenshot shows the AWS Management Console interface. On the left is the navigation pane with categories like 'Panel de EC2', 'Eventos', 'Instancias', 'Imágenes', and 'Elastic Block Store'. The main area displays the 'Instancias (1/1)' page. A table lists the instances, with 'Command Host' (ID: i-094b2f4425775602a) in the 'En ejecución' state. Below the table, the details for the selected instance are shown, including its public IP (44.242.224.3), private IP (10.0.1.92), and DNS name (ec2-44-242-224-3.us-west-2.compute.amazonaws.com).

Name	ID de la instancia	Estado de la instancia	Tipo de instancia	Comprobación de estado	Estado de la alarma	Zona de disponibilidad
Command Host	i-094b2f4425775602a	En ejecución	t3.medium	3/3 comprobador	Ver alarmas	us-west-2

i-094b2f4425775602a (Command Host)		
Resumen de instancia		
ID de la instancia i-094b2f4425775602a (Command Host)	Dirección IPv4 pública 44.242.224.3 dirección abierta	Direcciones IPv4 privadas 10.0.1.92
Dirección IPv6 -	Estado de la instancia En ejecución	DNS de IPv4 pública ec2-44-242-224-3.us-west-2.compute.amazonaws.com



- Seleccione Conectar.
- En la pestaña Conectar instancia EC2, seleccione Conectar.

El puerto 22 (SSH) se encuentra abierto a todas las direcciones IPv4.

El puerto 22 (SSH) actualmente se encuentra abierto a todas las direcciones IPv4, indicadas por 0.0.0.0/0 en la regla de entrada del [grupo de seguridad](#). Para aumentar la seguridad, considere la posibilidad de restringir el acceso únicamente a las direcciones IP del servicio Conexión de instancias de EC2 para la región: 18.237.140.160/29. [Más información](#).

ID de la instancia
i-094b2f4425775602a (Command Host)

Tipo de conexión

☒ Conectarse mediante la Conexión de la instancia EC2
Conéctese mediante el cliente basado en navegador de EC2 Instance Connect, con una dirección IPv4 pública.

☐ Conectarse mediante punto de conexión de EC2 Instance Connect
Conéctese mediante el cliente basado en navegador de EC2 Instance Connect, con una dirección IPv4 privada y un punto de conexión de VPC.

Dirección IP pública
44.242.224.3

Nombre de usuario
Escriba el nombre de usuario definido en la AMI utilizada para lanzar la instancia. Si no definió un nombre de usuario personalizado, utilice el nombre de usuario predeterminado, ec2-user.

ec2-user

Nota: En la mayoría de los casos, el nombre de usuario predeterminado, ec2-user, es correcto. Sin embargo, lea las instrucciones de uso de la AMI para comprobar si el propietario de la AMI ha cambiado el nombre de usuario predeterminado.

Cancelar Conectar

- Para confirmar que la región en la que se ejecuta la instancia del host de comando es la misma que la del laboratorio (la región us-west-2), ejecute el siguiente comando:

```
[ec2-user@ip-10-0-1-92 ~]$ curl http://169.254.169.254/latest/dynamic/instance-identity/document | grep region
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  474    100  474    0    0   263k      0  --:--:-- --:--:-- --:--:--   462k
"region" : "us-west-2",
[ec2-user@ip-10-0-1-92 ~]$
```

- Para actualizar el software AWS CLI con las credenciales correctas, ejecute aws configure:
- En las indicaciones, ingrese la siguiente información:

```
[ec2-user@ip-10-0-1-92 ~]$ aws configure
AWS Access Key ID [None]:
AWS Secret Access Key [None]:
Default region name [us-west-2]:
Default output format [None]: json
[ec2-user@ip-10-0-1-92 ~]$
```

- Para acceder a estos scripts, ingrese cd /home/ec2-user :



- Para inspeccionar el script UserData.txt que se instaló como parte de la creación del host de comando, ejecute el siguiente comando:

```
[ec2-user@ip-10-0-1-92 ~]$ more UserData.txt
#!/bin/bash
yum update -y --security
amazon-linux-extras install epel -y
yum -y install httpd php stress
systemctl enable httpd.service
systemctl start httpd
cd /var/www/html
wget http://aws-tc-largeobjects.s3.amazonaws.com/CUR-TF-100-TULABS-1/10-lab-autoscaling-linux/s3/ec2-stress.zip
unzip ec2-stress.zip

echo 'UserData has been successfully executed. ' >> /home/ec2-user/result
find -wholename /root/.*history -wholename /home/*.*history -exec rm -f {} \;
find / -name 'authorized_keys' -exec rm -f {} \;
rm -rf /var/lib/cloud/data/scripts/*
[ec2-user@ip-10-0-1-92 ~]$
```

- En la parte superior de esta página, seleccione Detalles y seleccione Mostrar.
- Copie los valores KEYNAME, AMIID, HTTPACCESS y SUBNETID en un documento de editor de texto y luego elija X para cerrar el panel Credenciales.

AMIID	ami-0f6cac0240f22d17e
HTTPACCESS	sg-01caf6c48b144ecc7
COMMANDHOSTIP	44.242.224.3
KEYNAME	vockey
SUBNETID	subnet-048da4ba9596b224c

- En el siguiente script, reemplace el texto correspondiente con los valores del paso anterior.
- Ingrese el script modificado en la ventana del terminal y ejecute el script.

```
[ec2-user@ip-10-0-1-92 ~]$ aws ec2 run-instances --key-name vockey --instance-type t3.micro --image-id ami-0f6cac0240f22d17e --user-data file:///home/ec2-user/UserData.txt --security-group-ids sg-01caf6c48b144ecc7 --subnet-id subnet-048da4ba9596b224c --associate-public-ip-address - --tag-specifications 'ResourceType=instance,Tags=[{Key=Name,Value=WebServer}]' --output text --query 'Instances[*].InstanceId'
i-0bb473a30ae04c507
[ec2-user@ip-10-0-1-92 ~]$
```

- Copie y pegue el valor InstanceId en un editor de texto para usarlo más tarde.

i-0bb473a30ae04c507



- o Para utilizar el comando `aws ec2 wait instance-running` para supervisar el estado de esta instancia, reemplace `NEW-INSTANCE-ID` en el siguiente comando con el valor `InstanceID` que copió en el paso anterior. Ejecute el comando modificado.

```
[ec2-user@ip-10-0-1-92 ~]$ aws ec2 wait instance-running --instance-ids i-0bb473a30ae04c507
[ec2-user@ip-10-0-1-92 ~]$
```

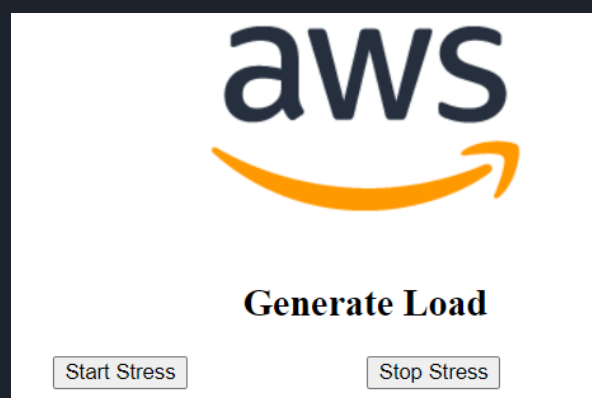
- o Para obtener el nombre DNS público, en el siguiente comando, reemplace `NEW-INSTANCE-ID` con el valor que copió anteriormente y ejecute el comando modificado:

```
[ec2-user@ip-10-0-1-92 ~]$ aws ec2 describe-instances --instance-id i-0bb473a30ae04c507 --query 'Reservations[0].Instances[0].NetworkInterfaces[0].Association.PublicDnsName'
"ec2-54-245-15-28.us-west-2.compute.amazonaws.com"
[ec2-user@ip-10-0-1-92 ~]$
```

- o Copie la salida de este comando sin las comillas.

```
ec2-54-245-15-28.us-west-2.compute.amazonaws.com
```

- o En una nueva pestaña del navegador, ingrese la salida que copió del paso anterior.
- o En el siguiente comando, reemplace `PUBLIC-DNS-ADDRESS` con el valor que copió en los pasos anteriores y luego ejecute el comando modificado.



- o Para crear una nueva AMI basada en esta instancia, en el siguiente comando `aws ec2 create-image`, reemplace `NEW-`



INSTANCE-ID con el valor que copió anteriormente y ejecute el comando ajustado:

```
[ec2-user@ip-10-0-1-92 ~]$ aws ec2 create-image --name WebServerAMI --instance-id i-0bb473a30ae04c507  
{  
  "ImageId": "ami-01104f6e7f74fecc2"  
}  
[ec2-user@ip-10-0-1-92 ~]$
```



TAREA 2

En esta sección, creará un balanceador de carga que agrupe un grupo de instancias de EC2 bajo una única dirección del Sistema de nombres de dominio (DNS). Utilizará el escalado automático para crear un grupo de instancias de EC2 escalable de forma dinámica en función de la imagen que creó en la tarea anterior. Por último, creará un conjunto de alarmas que aumenten o reduzcan la cantidad de instancias en su grupo de balanceadores de carga siempre que el rendimiento de la CPU de cualquier máquina dentro del grupo supere o descienda por debajo de un conjunto de umbrales especificados.

- En la Consola de administración de EC2, en el panel de navegación izquierdo, ubique la sección Equilibrio de carga y seleccione Equilibradores de carga.
- Seleccione Crear balanceador de carga.
- En la sección Tipos de balanceador de carga, para Balanceador de carga de aplicaciones, seleccione Crear.



Balancedador de carga de aplicaciones [Info](#)

Elija un balanceador de carga de aplicaciones cuando necesite un conjunto de características flexibles para sus aplicaciones con tráfico HTTP y HTTPS. En el nivel de solicitud, los balanceadores de carga de aplicaciones proporcionan características avanzadas de enrutamiento y visibilidad dirigidas a arquitecturas de aplicación, incluidos microservicios y contenedores.

[Crear](#)

Balancedador de carga de red [Info](#)

Elija un equilibrador de carga de red cuando necesite un rendimiento ultraalto, descarga de TLS a gran escala, implementación centralizada de certificados, compatibilidad con UDP y direcciones IP estáticas para sus aplicaciones. En el nivel de conexión, los equilibradores de carga de red pueden controlar millones de solicitudes por segundo de forma segura a la vez que mantienen latencias ultrabajas.

[Crear](#)

Equilibrador de carga de gateway [Info](#)

Elija un equilibrador de carga de gateway cuando necesite implementar y administrar una flota de dispositivos virtuales de terceros compatibles con GENEVE. Estos dispositivos permiten mejorar los controles de las políticas, la seguridad y la conformidad.

[Crear](#)

- o En la página Crear balanceador de carga de aplicaciones, en la sección Configuración básica, configure la siguiente opción:

Configuración básica

Nombre del balanceador de carga
Debe ser nombre único dentro de su cuenta de AWS y no puede cambiarse después de crear el equilibrador de carga.

Se permite un máximo de 32 caracteres alfanuméricos, incluidos guiones, pero el nombre no puede comenzar ni terminar por un guión.

- o En la sección Mapeo de red, configure las siguientes opciones para las dos zonas de disponibilidad:



Mapeo de red [Info](#)

El balanceador de carga dirige el tráfico a los destinos de las subredes seleccionadas y en función de la configuración de las direcciones IP.

VPC [Info](#)

The load balancer will exist and scale within the selected VPC. The selected VPC is also where the load balancer targets must be hosted unless routing to Lambda or on-premises targets, or if using VPC peering. To confirm the VPC for your targets, view [target groups](#). For a new VPC, [create a VPC](#).

Lab VPC
vpc-09a2689aacf4d4beaf
CIDR de VPC IPv4: 10.0.0.0/16

Mapeos [Info](#)

Seleccione al menos dos zonas de disponibilidad y una subred por zona. El equilibrador de carga solo dirige el tráfico a los destinos de estas zonas de disponibilidad. Las zonas de disponibilidad que no son compatibles con el equilibrador de carga o la VPC no están disponibles para seleccionarlás.

Zonas de disponibilidad

☒ **us-west-2a (usw2-az1)**

Subred

subnet-0634c0ae124489f7a
IPv4 subnet CIDR: 10.0.1.0/24

Public Subnet 1

Dirección IPv4
Asignado por AWS

☒ **us-west-2b (usw2-az2)**

Subred

subnet-048da4ba9596b224c
IPv4 subnet CIDR: 10.0.3.0/24

Public Subnet 2

Dirección IPv4
Asignado por AWS

- En la sección Grupos de seguridad, elija la X del grupo de seguridad predeterminado para eliminarlo.
- En la lista desplegable Grupos de seguridad, elija HTTPAccess.

Grupos de seguridad [Info](#)

Un grupo de seguridad consiste en un conjunto de reglas de firewall que controlan el tráfico hacia el equilibrador de carga. Seleccione un grupo de seguridad existente o [cree un nuevo grupo de seguridad](#).

Grupos de seguridad

Seleccione hasta 5 grupos de seguridad

HTTPAccess
sg-01caf6c48b144ecc7 VPC: vpc-09a2689aacf4d4beaf

- En la sección Oyentes y enrutamiento, seleccione el enlace Crear grupo de destino.
- En la página Especificar detalles del grupo, en la sección Configuración básica, configure las siguientes opciones:



Configuración básica

La configuración de esta sección no se puede cambiar después de crear el grupo de destino.

Elegir un tipo de destino

☒ Instancias

- Admite el balanceo de carga en instancias dentro de una VPC específica.
- Facilita el uso de [Amazon EC2 Auto Scaling](#) para administrar y escalar la capacidad de EC2.

☐ Direcciones IP

- Admite el balanceo de carga en recursos de VPC y en las instalaciones.
- Facilita el direccionamiento a varias direcciones IP e interfaces de red en la misma instancia.
- Ofrece flexibilidad con arquitecturas basadas en microservicios, lo que simplifica la comunicación entre aplicaciones.
- Admite destinos IPv6, lo que permite la comunicación IPv6 integral y NAT de IPv4 a IPv6.

☐ Función Lambda

- Facilita el direccionamiento a una única función Lambda.
- Accesible solo para balanceadores de carga de aplicaciones.

☐ Balanceador de carga de aplicaciones

- Ofrece la flexibilidad para que un balanceador de carga de red acepte y dirija solicitudes TCP dentro de una VPC específica.
- Facilita el uso de direcciones IP estáticas y PrivateLink con un balanceador de carga de aplicaciones.

Nombre del grupo de destino

webserver-app

Se permite un máximo de 32 caracteres alfanuméricos, incluidos guiones, pero el nombre no puede comenzar ni terminar por un guion.

- o En la sección Comprobaciones de salud, en Ruta de comprobación de salud, ingrese/index.php

Comprobaciones de estado

El balanceador de carga asociado envía periódicamente solicitudes, según las configuraciones que aparecen a continuación, a los destinos registrados para comprobar su estado.

Protocolo de comprobación de estado

HTTP

Ruta de comprobación de estado

Utilice la ruta predeterminada "/" para realizar comprobaciones de estado en la raíz, o especifique una ruta personalizada si lo prefiere.

/index.php

Hasta 1024 caracteres permitidos.

► Configuración avanzada de comprobación de estado

- o En la parte inferior de la página, seleccione Siguiente.
- o En la página Registrar objetivos, seleccione Crear grupo de objetivos.



Se creó correctamente el grupo de destinos: **webserver-app**. La detección de anomalías se aplica automáticamente a todos los destinos registrados. El estado se puede ver en la pestaña **Destinos**.

EC2 > Grupos de destino > webserver-app

webserver-app

Acciones ▼

Detalles

arn:aws:elasticloadbalancing:us-west-2:867962060815:targetgroup/webserver-app/841259a143874c42

Tipo de destino Instancia	Protocolo : Puerto HTTP: 80	Versión del protocolo HTTP1	VPC vpc-09a2689aacf4dbef
Tipo de dirección IP IPv4	Balancedador de carga Ninguno asociado		

0 Destinos totales	0 En buen estado	0 En mal estado	0 Sin utilizar	0 Inicial	0 Vaciado
	0 Anómalo				

- Regrese a la pestaña del navegador Balanceadores de carga y localice la sección Escuchas y enrutamiento. Para Acción predeterminada, elija Actualice a la derecha de la lista desplegable Reenviar a.
- En la lista desplegable Reenviar a, elija webserver-app.

Agentes de escucha y direccionamiento Info

Un agente de escucha es un proceso que comprueba las solicitudes de conexión mediante el puerto y el protocolo que configure. Las reglas que defina para un agente de escucha determinan cómo el equilibrador de carga dirige las solicitudes a sus destinos registrados.

▼ Agente de escucha HTTP:80 Eliminar

Protocolo HTTP ▼	Puerto 80 1-65535	Acción predeterminada <small>Info</small> Reenviar a webserver-app Tipo de destino: Instancia, IPv4 Crear un grupo de destino	HTTP ▼	
---------------------	-------------------------	--	--------	--

Etiquetas del agente de escucha - *opcional*
Considere la posibilidad de agregar etiquetas al agente de escucha. Las etiquetas permiten clasificar los recursos de AWS para que pueda administrarlos con mayor facilidad.

Agregar etiqueta de agente de escucha
Puede agregar hasta 50 etiquetas más.

Agregar agente de escucha

- En la parte inferior de la página, seleccione Crear balanceador de carga.



Se creó correctamente el equilibrador de carga: **WebServerELB**

Pueden transcurrir unos minutos hasta que el equilibrador de carga esté totalmente configurado y listo para dirigir el tráfico. Los destinos también tardarán unos minutos en completar el proceso de registro y superar las comprobaciones de estado iniciales.

EC2 > Balanceadores de carga > WebServerELB

WebServerELB

Acciones ▼

▼ Detalles

Tipo de equilibrador de carga	Estado	VPC	Tipo de dirección IP del equilibrador de carga
Aplicación	Aprovisionándose	vpc-09a2689aacf4dbeeaf	IPv4
Esquema	Zona hospedada	Zonas de disponibilidad	Fecha creada
Internet-facing	Z1H1FL5HAB5F5	subnet-048da4ba9596b224c us-west-2b (usw2-az2)	28 de septiembre de 2024, 23:16 (UTC-05:00)
		subnet-0634c0ae124489f7a us-west-2a (usw2-az1)	
ARN del equilibrador de carga		Nombre de DNS info	
<code>arn:aws:elasticloadbalancing:us-west-2:867962060815:loadbalance/r/app/WebServerELB/103aabd528aa8df3</code>		<code>WebServerELB-465551264.us-west-2.elb.amazonaws.com</code> (Registro A)	

- Para ver el balanceador de carga WebServerELB que creó, seleccione Ver balanceador de carga.
- Para copiar el nombre DNS del balanceador de carga, utilice la opción de copiar y pegue el nombre DNS en un editor de texto.

DNS ELB: `WebServerELB-465551264.us-west-2.elb.amazonaws.com`

- En la Consola de administración de EC2, en el panel de navegación izquierdo, ubique la sección Instancias y elija Plantillas de lanzamiento.
- Seleccione Crear plantilla de lanzamiento.
- En la página Crear plantilla de lanzamiento, en la sección Nombre y descripción de la plantilla de lanzamiento, configure las siguientes opciones:



Crear plantilla de lanzamiento

La creación de una plantilla de lanzamiento le permite crear una configuración de instancia guardada que se puede reutilizar, compartir y lanzar más adelante. Las plantillas pueden tener varias versiones.

Nombre y descripción de la plantilla de lanzamiento

Nombre de la plantilla de lanzamiento - *obligatorio*

Debe ser única para esta cuenta. Máximo de 128 caracteres. Sin espacios ni caracteres especiales, como "&", "*", "@".

Descripción de la versión de la plantilla

Máximo de 255 caracteres

Orientación sobre Auto Scaling | [Información](#)

Seleccíonelo si va a utilizar esta plantilla con EC2 Auto Scaling

☒ Proporcionar orientación que me ayude a configurar una plantilla que pueda utilizar con EC2 Auto Scaling

► Etiquetas de la plantilla

► Plantilla de origen

- En la sección Imágenes de aplicaciones y sistemas operativos (imagen de máquina de Amazon): requeridas, elija la pestaña Mis AMI.

▼ Imágenes de aplicaciones y sistemas operativos (Imagen de máquina de Amazon) - obligatorio [Información](#)

Una AMI es una plantilla que contiene la configuración de software (sistema operativo, servidor de aplicaciones y aplicaciones) necesaria para lanzar la instancia. Busque o examine las AMI si no ve lo que busca a continuación.

Recientes

Mis AMI

Inicio rápido

☒ De mi propiedad

☐ Compartido conmigo

Buscar más AMI

Inclusión de AMI de AWS, Marketplace y la comunidad

Imágenes de máquina de Amazon (AMI)

WebServerAMI

ami-01104f6e7f74fecc2

2024-09-29T03:59:15.000Z

Virtualización: hvm

Activado para ENA: true

Tipo de dispositivo raíz: ebs

Descripción

-

Arquitectura

ID de AMI

x86_64

ami-01104f6e7f74fecc2

- En la sección Tipo de instancia, elija la lista desplegable Tipo de instancia y elija t3.micro.



▼ **Tipo de instancia** [Información](#) | [Obtener asesoramiento](#) Avanzado

Tipo de instancia

t3.micro
Familia: t3 2 vCPU 1 GiB Memoria Generación actual: true
Bajo demanda SUSE base precios: 0.0104 USD por hora
Bajo demanda Windows base precios: 0.0196 USD por hora
Bajo demanda RHEL base precios: 0.0392 USD por hora
Bajo demanda Linux base precios: 0.0104 USD por hora

[Se aplican costos adicionales a las AMI con software preinstalado](#)

☒ Todas las generaciones
[Comparar tipos de instancias](#)

- En la sección Par de claves (inicio de sesión), confirme que la lista desplegable Nombre del par de claves esté configurada en No incluir en la plantilla de inicio.

▼ **Par de claves (inicio de sesión)** [Información](#)

Puede utilizar un par de claves para conectarse de forma segura a la instancia. Asegúrese de que tiene acceso al par de claves seleccionado antes de lanzar la instancia.

Nombre del par de claves

No incluir en la plantilla de lanzamiento ▼

[Crear un nuevo par de claves](#)

- En la sección Configuración de red, elija la lista desplegable Grupos de seguridad y elija HTTPAccess

▼ **Configuraciones de red** [Información](#)

Subred [Información](#)

No incluir en la plantilla de lanzamiento ▼

[Crear nueva subred](#)

Al especificar una subred, se agrega automáticamente una interfaz de red a la plantilla.

Firewall (grupos de seguridad) [Información](#)

Un grupo de seguridad es un conjunto de reglas de firewall que controlan el tráfico de la instancia. Agregue reglas para permitir que un tráfico específico llegue a la instancia.

☒ Seleccionar un grupo de seguridad existente

☐ Crear grupo de seguridad

Grupos de seguridad [Información](#)

Seleccionar grupos de seguridad ▼

[Compare reglas de grupo de seguridad](#)

HTTPAccess sg-01caf6c48b144ecc7 ✕

VPC: vpc-09a2689aacf4dbeaf

▶ Configuración de red avanzada

- Seleccione Crear plantilla de lanzamiento
- Seleccione Ver plantillas de lanzamiento.



Plantillas de lanzamiento (1/1) Información

Buscar

ID de plantilla de lanzamiento	Nombre de plantilla de lanzamiento	Versión predeterminada	Última versión	Hora de creación
lt-0f950d6aba82f968a	web-app-launch-template	1	1	2024-09-29T04:28:5

web-app-launch-template (lt-0f950d6aba82f968a)

Detalles de la plantilla de lanzamiento

Acciones Eliminar plantilla

ID de la plantilla de lanzamiento lt-0f950d6aba82f968a	Nombre de la plantilla de lanzamiento web-app-launch-template	Versión predeterminada 1	Propietario arn:aws:sts::867962060815:assumed-role/voclabs/user3386630=Joseph_Julios
---	--	-----------------------------	---

Detalles Versiones Etiquetas de la plantilla

- Elegir web-app-launch-template y luego desde Acciones lista desplegable, elija Crear grupo de escalamiento automático.
- En la página Elegir plantilla de lanzamiento o configuración, en la sección Nombre, para Nombre del grupo de Auto Scaling, ingrese Web App Auto Scaling Group

Nombre

Nombre del grupo de Auto Scaling

Escriba un nombre para identificar el grupo.

Web App Auto Scaling Group

Debe ser único para esta cuenta en la región actual y no puede superar los 255 caracteres.

- Seleccione Siguiente.
- En la página Elegir opciones de lanzamiento de instancia, en la sección Red, configure las siguientes opciones:



Red [Info](#)

Para la mayoría de las aplicaciones, puede utilizar varias zonas de disponibilidad y dejar que EC2 Auto Scaling equilibre sus instancias entre las zonas. La VPC predeterminada y las subredes predeterminadas son adecuadas para comenzar rápidamente.

VPC
Elija la VPC que define la red virtual para el grupo de Auto Scaling.

vpc-09a2689aacf4dbeaf (Lab VPC)
10.0.0.0/16

[Crear una VPC](#)

Zonas de disponibilidad y subredes
Defina qué zonas de disponibilidad y subredes puede utilizar el grupo de Auto Scaling en la VPC elegida.

Seleccionar zonas de disponibilidad y subredes

us-west-2a | subnet-0e0eb16a2fda00906 (Private Subnet 1)
10.0.2.0/24

us-west-2b | subnet-07a553c18ee72be22 (Private Subnet 2)
10.0.4.0/24

[Crear una subred](#)

- Seleccione Siguiente.
- En la página Configurar opciones avanzadas – opcional, configure las siguientes opciones:

Balance de carga [Info](#)

Utilice las siguientes opciones para asociar su grupo de Auto Scaling a un balanceador de carga existente o a uno nuevo que defina.

☐ No se encontró ningún balanceador de carga
El tráfico a su grupo de Auto Scaling no se llevará a cabo por un balanceador de carga.

☒ Asociar a un balanceador de carga existente
Elija entre los balanceadores de carga existentes.

☐ Asociar a un nuevo balanceador de carga
Cree rápidamente un balanceador de carga básico para asociarlo al grupo de Auto Scaling.

Asociar a un balanceador de carga existente
Seleccione los balanceadores de carga que desea asociar al grupo de Auto Scaling.

☒ Elegir entre los grupos de destino del balanceador de carga
Esta opción le permite asociar balanceadores de carga de puerta de enlace, red o aplicaciones.

☐ Elegir entre balanceadores de carga clásicos

Grupos de destino del balanceador de carga existentes
Solo están disponibles para su selección los grupos de destino de instancias que pertenecen a la misma VPC que el grupo de Auto Scaling.

Seleccionar grupos de destino

webserver-app | HTTP
Application Load Balancer: WebServerELB

- Seleccione Siguiente.
- En la página Configurar políticas de escala y tamaño de grupo (opcional), configure las siguientes opciones:



Límites de escalamiento
Establezca límites sobre cuánto puede aumentarse o disminuirse la capacidad deseada.

Capacidad deseada mínima	Capacidad deseada máxima
<input type="text" value="2"/>	<input type="text" value="4"/>
Capacidad igual o inferior a la deseada	Capacidad igual o superior a la deseada

Escalamiento automático - opcional
Elija si desea utilizar una política de seguimiento de destino [Info](#)
Puede configurar otras políticas de escalado basadas en métricas y un escalado programado después de crear su grupo de escalamiento automático.

☐ Sin políticas de escalamiento
Su grupo de escalamiento automático mantendrá su tamaño inicial y no se redimensionará de forma dinámica para satisfacer la demanda.

☒ Política de escalado de seguimiento de destino
Elija una métrica y un valor objetivo de CloudWatch y deje que la política de escalamiento ajuste la capacidad deseada en proporción al valor de la métrica.

Nombre de la política de escalado

Tipo de métrica [Info](#)
Métrica supervisada que determina si la utilización de recursos es demasiado baja o alta. Si utiliza métricas de EC2, considere la posibilidad de habilitar la supervisión detallada para obtener un mejor rendimiento de escalado.

Valor de destino

Preparación de la instancia [Info](#)
 segundos

☐ Deshabilite el escalado descendente para crear solo una política de escalado ascendente

- Seleccione Siguiente.
- En la página Agregar notificaciones – opcional, seleccione Siguiente.
- En la página Agregar etiquetas (opcional), elija Agregar etiqueta y configure las siguientes opciones:

Etiquetas (1)

Clave	Valor: opcional	Etiquetar instancias nuevas	
<input type="text" value="Name"/>	<input type="text" value="WebApp"/>	<input checked="" type="checkbox"/>	<input type="button" value="Eliminar"/>

49 restante

- Seleccione Siguiente.
- En la página Revisar, seleccione Crear grupo de escalamiento automático.



Web App Auto Scaling Group, 1 Política de escalado se creó correctamente

EC2 > Grupos de Auto Scaling

Grupos de Auto Scaling (1) Info

Configuraciones de lanzamiento

Plantillas de lanzamiento

Acciones

Crear grupo de escalado automático

Buscar sus grupos de Auto Scaling

< 1 > ⚙

<input type="checkbox"/>	Nombre	Plantilla de lanzamiento/config...	Instanc...	Estado	Capa
<input type="checkbox"/>	Web App Auto Scaling Group	web-app-launch-template Versión Prede	0	⌵ Actualización de cap	2



TAREA 3

En esta tarea, verifica que tanto la configuración de escalamiento automático como el balanceador de carga estén funcionando accediendo a un script preinstalado en uno de tus servidores que consumirá ciclos de CPU, lo que invoca la alarma de escalamiento horizontal.

- En el panel de navegación izquierdo, seleccione Instancias.

Instancias (4) Información							
Última actualización Hace 1 minute		Conectar		Estado de la instancia ▼		Acciones ▼	
Lanzar instancias ▼							
Buscar Instancia por atributo o etiqueta (case-sensitive) Todos los estados ▼							
<input type="checkbox"/>	Name	ID de la instancia	Estado de la i...	Tipo de inst...	Comprobación de	Estado de la al...	Zona
<input type="checkbox"/>	Command Host	i-094b2f4425775602a	En ejecución	t3.medium	3/3 comprobador	Ver alarmas +	us-we
<input type="checkbox"/>	WebServer	i-0bb473a30ae04c507	En ejecución	t3.micro	3/3 comprobador	Ver alarmas +	us-we
<input type="checkbox"/>	WebApp	i-05a4fe2da38016bd6	En ejecución	t3.micro	2/2 comprobador	Ver alarmas +	us-we
<input type="checkbox"/>	WebApp	i-0d3ea0ef66e81eb64	En ejecución	t3.micro	3/3 comprobador	Ver alarmas +	us-we

- Una vez que las instancias hayan completado la inicialización, en el panel de navegación izquierdo en la sección Equilibrio de carga, elija Grupos de destino y luego seleccione Su grupo objetivo, aplicación de servidor web.
- En la pestaña Destinos, verifique que se estén creando dos instancias. Actualice esta lista hasta que el estado de estas instancias cambie a saludable.



[EC2](#) > Grupos de destino

Grupos de destino (1/1) [Info](#)

Acciones ▾

Crear un grupo de destino

Filtrar grupos de destino

< 1 > ⚙

<input checked="" type="checkbox"/>	Nombre ▾	ARN ▾	Puerto ▾	Protocolo ▾	Tipo de destino ▾
<input checked="" type="checkbox"/>	webserver-app	arn:aws:elasticloadbalanci...	80	HTTP	Instancia

Grupo de destino: webserver-app

Detalles

Destinos

Monitorización

Comprobaciones de estado

Atributos

Etiquetas

Destinos registrados (2) [Info](#)

Mitigación de anomalías:No aplicable

Anular el registro

Registrar destinos

Los grupos de destinos enrutan las solicitudes a destinos individuales registrados mediante el protocolo y el número de puerto que especifique. Las comprobaciones de estado se realizan en todos los destinos registrados de acuerdo con la configuración de comprobación de estado del grupo de destinos. La detección de anomalías se aplica automáticamente a los grupos de destinos de HTTP/HTTPS con al menos 3 destinos en buen estado.

Filtrar destinos

< 1 > ⚙

<input type="checkbox"/>	ID de instancia ▾	Nombre ▾	Puerto ▾	Zona ▾	Estado ▾	Detalles del estado
<input type="checkbox"/>	i-05a4fe2da38016bd6	WebApp	80	us-west-2b	Healthy	-
<input type="checkbox"/>	i-0d3ea0ef66e81eb64	WebApp	80	us-west-2a	Healthy	-

aws

re/start



TAREA 4

Prueba de la configuración de escalado automático.

- Abra una nueva pestaña del navegador web, pegue el nombre DNS del balanceador de carga que copió anteriormente en la barra de direcciones y presione Entrar.
- En la página web, seleccione Iniciar estrés.



- En la consola de administración de EC2, en el panel de navegación izquierdo de la sección Escalado automático, seleccione Grupos de escalado automático.
- Seleccionar Grupo de escalamiento automático de aplicaciones web.
- Seleccione la pestaña Actividad.

Grupo de Auto Scaling: Web App Auto Scaling Group		
Estado	Descripción	Causa
Correcto	Launching a new EC2 instance: i-05a4fe2da38016bd6	At 2024-09-29T04:37:59Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 2. At 2024-09-29T04:38:05Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 2.
Correcto	Launching a new EC2 instance: i-0d3ea0ef66e81eb64	At 2024-09-29T04:37:59Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 2. At 2024-09-29T04:38:05Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 2.